PLS post processing by similarity transformation (PLS+ST): a simple alternative to OPLS Theoretical properties and proofs

Rolf Ergon
Telemark University College Porsgrunn, Norway

Email: rolf.ergon@hit.no

This Supplementary Appendix gives the details and proofs of properties and results in the paper *PLS post-processing by similarity transformation (PLS+ST): A simple alternative to OPLS* [9]. For the readers convenience, the OPLS algorithm [2] is also included.

Property 1 The Martens factorization (2) has the special property that all score vectors except the first one are orthogonal to both $\mathbf{y}$ and $\hat{\mathbf{y}}$.

Proof: Since $\mathbf{w}_1$ is given by $\mathbf{w}_1 = \frac{\mathbf{X}^T\mathbf{y}}{\|\mathbf{X}^T\mathbf{y}\|}$ and $\mathbf{T}_{2:A} = \mathbf{X}\mathbf{W}_{2:A}$, and since $\mathbf{W}^T\mathbf{W} = \mathbf{I}$, it follows that $\mathbf{T}_{2:A}^T\mathbf{y} = \mathbf{W}_{2:A}^T\mathbf{X}^T\mathbf{y} = \|\mathbf{X}^T\mathbf{y}\|\,\mathbf{W}_{2:A}^T\mathbf{w}_1 = \mathbf{0}$. From the prediction formula (3) further follows $\hat{\mathbf{y}} = \mathbf{X}\mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y}$ and thus

$$\begin{aligned}\left[\begin{array}{cc}\mathbf{t}_1 & \mathbf{T}_{2:A}\end{array}\right]^T\hat{\mathbf{y}} &= \mathbf{T}^T\hat{\mathbf{y}} = \mathbf{W}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y} = \mathbf{W}^T\mathbf{X}^T\mathbf{y} \\ &= \|\mathbf{X}^\mathbf{T}\mathbf{y}\|\,\mathbf{W}^T\mathbf{w}_1 = \left[\begin{array}{cc}\|\mathbf{X}^\mathbf{T}\mathbf{y}\| & \mathbf{0}\end{array}\right]^T, \end{aligned} \tag{17}$$

i.e. $\mathbf{T}_{2:A}^T\hat{\mathbf{y}} = \mathbf{0}$.

Property 2 The residual $\mathbf{E}$ in the Martens factorization (2) is also orthogonal to $\mathbf{y}$.

Proof: From $\mathbf{X}^T\mathbf{y} = \|\mathbf{X}^T\mathbf{y}\|\,\mathbf{w}_1$, $\mathbf{w}_1^T\mathbf{w}_1 = 1$ and $\mathbf{y}^T\mathbf{T}_{2:A} = 0$ follows $\mathbf{y}^T\mathbf{E} = \mathbf{y}^T\left(\mathbf{X} - \mathbf{T}\mathbf{W}^T\right) = \mathbf{y}^T\mathbf{X} - \mathbf{y}^T\left(\mathbf{t}_1\mathbf{w}_1^T + \mathbf{T}_{2:A}\mathbf{W}_{2:A}^T\right) = \mathbf{y}^T\mathbf{X} - \mathbf{y}^T\mathbf{t}_1\mathbf{w}_1^T = \mathbf{y}^T\mathbf{X} - \mathbf{y}^T\mathbf{X}\mathbf{w}_1\mathbf{w}_1^T = \|\mathbf{X}^T\mathbf{y}\|\,\mathbf{w}_1^T - \|\mathbf{X}^T\mathbf{y}\|\,\mathbf{w}_1^T = \mathbf{0}$.

Property 3 The factorizations (13) and (2) are identical, i.e. $\mathbf{T}_\mathrm{W}\mathbf{P}^T\mathbf{W} = \mathbf{T}$.

Proof: From the two well known estimator expressions $\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y}$ and $\hat{\mathbf{b}} = \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\mathbf{q}_\mathrm{W} = \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\left(\mathbf{T}_\mathrm{W}^T\mathbf{T}_\mathrm{W}\right)^{-1}\mathbf{T}_\mathrm{W}^T\mathbf{y}$ [7] follows

$$\mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y} = \mathbf{W}\left(\left(\mathbf{P}^T\mathbf{W}\right)^T\mathbf{T}_\mathrm{W}^T\mathbf{T}_\mathrm{W}\mathbf{P}^T\mathbf{W}\right)^{-1}\left(\mathbf{P}^T\mathbf{W}\right)^T\mathbf{T}_\mathrm{W}^T\mathbf{y}, \tag{18}$$

i.e. $\mathbf{T}_\mathrm{W}\mathbf{P}^T\mathbf{W} = \mathbf{X}\mathbf{W} = \mathbf{T}$.

Property 4 The loading matrices in the factorizations (12) and (13) are $\mathbf{P} = \left[\begin{array}{ccccc}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_{A-1} & \mathbf{p}_A\end{array}\right]$ and $\mathbf{W}\mathbf{W}^T\mathbf{P} = \left[\begin{array}{ccccc}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_{A-1} & \mathbf{w}_A\end{array}\right]$, i.e. they are different in the last column vector only.

Proof: The orthogonalized PLS algorithm results in an upper triangular and bi-diagonal matrix $\mathbf{P}^T\mathbf{W}$, with ones along the main diagonal [8]. We thus have

$$\mathbf{P}^T\mathbf{W}\mathbf{W}^T = \begin{bmatrix} 1 & \mathbf{p}_1^T\mathbf{w}_2 & 0 & \cdots & 0 \\ 0 & 1 & \mathbf{p}_2^T\mathbf{w}_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 & \mathbf{p}_{A-1}^T\mathbf{w}_A \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \vdots \\ \mathbf{w}_A^T \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^T + \mathbf{p}_1^T\mathbf{w}_2\mathbf{w}_2^T \\ \mathbf{w}_2^T + \mathbf{p}_2^T\mathbf{w}_3\mathbf{w}_3^T \\ \vdots \\ \mathbf{w}_{A-1}^T + \mathbf{p}_{A-1}^T\mathbf{w}_A\mathbf{w}_A^T \\ \mathbf{w}_A^T \end{bmatrix}. \quad (19)$$

From this follows that $\mathbf{p}_A$ in $\mathbf{P}$ is replaced by $\mathbf{w}_A$, as stated. For a complete proof we must also show that for $2 \le i \le A$ we have $\mathbf{w}_{i-1}^T + \mathbf{p}_{i-1}^T\mathbf{w}_i\mathbf{w}_i^T = \mathbf{p}_{i-1}^T$, or equivalently that $\mathbf{w}_i^T = \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)$. Forming $\left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j$ we find the following possibilities for $2 \le i \le A$:

$$\begin{array}{rcl} j < i-1 & \Rightarrow & \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j = \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}(0-0) = 0 \\ j = i-1 & \Rightarrow & \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j = \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}(1-1) = 0 \\ j = i & \Rightarrow & \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j = \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T\mathbf{w}_i - 0\right) = 1 \\ j > i & \Rightarrow & \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j = \left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}(0-0) = 0 \end{array} \quad (20)$$

Since $\mathbf{p}_i$ and thus also $\left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1} - \mathbf{w}_{i-1}\right)$ belong to the span of $\mathbf{w}_1$, $\mathbf{w}_2$, ... $\mathbf{w}_A$, and since $\left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right)\mathbf{w}_j = 1$ for $j = i$ and 0 for $j \neq i$, it finally follows from the orthonormality of $\mathbf{W}$ that $\left(\mathbf{p}_{i-1}^T\mathbf{w}_i\right)^{-1}\left(\mathbf{p}_{i-1}^T - \mathbf{w}_{i-1}^T\right) = \mathbf{w}_i^T$, and thus that $\mathbf{w}_{i-1}^T + \mathbf{p}_{i-1}^T\mathbf{w}_i\mathbf{w}_i^T = \mathbf{p}_{i-1}^T$ for $2 \le i \le A$.

Property 5 Using a predetermined loading weights matrix $\mathbf{W}$, the deflation order in the algorithm resulting in the non-orthogonlized factorization (2) is of no importance for the residual and predictions. A loading weights matrix $\tilde{\mathbf{W}}$ with permuted column vectors will thus give $\mathbf{X} = \tilde{\mathbf{T}}\tilde{\mathbf{W}}^T + \mathbf{E}$ with $\tilde{\mathbf{T}}\tilde{\mathbf{W}}^T = \mathbf{T}\mathbf{W}^T$, and $y_{\text{new}} = \mathbf{x}_{\text{new}}^T\hat{\mathbf{b}}$ according to Eq. (3).

Proof: Since $\mathbf{W}$ is orthonormal the PLS algorithm giving the non-orthogonalized factorization (2) generally gives $\mathbf{t}_i = (\mathbf{X} - \sum_{\text{over all } j\neq i}\mathbf{t}_j\mathbf{w}_j^T)\mathbf{w}_i = \mathbf{X}\mathbf{w}_i$. This is true irrespective of the order of deflation, i.e. $\tilde{\mathbf{T}} = \mathbf{X}\tilde{\mathbf{W}}$. Introducing an invertible permutation matrix $\tilde{\mathbf{P}}$ with the property $\tilde{\mathbf{P}}^{-1} = \tilde{\mathbf{P}}^T$ and $\tilde{\mathbf{W}} = \mathbf{W}\tilde{\mathbf{P}}$, the predictions according to Eq. (3) will be $y_{\text{new}} = \mathbf{x}_{\text{new}}^T\tilde{\mathbf{b}} = \mathbf{x}_{\text{new}}^T\mathbf{W}\tilde{\mathbf{P}}\left(\tilde{\mathbf{P}}\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\tilde{\mathbf{P}}\right)^{-1}\tilde{\mathbf{P}}\mathbf{W}^T\mathbf{X}^T\mathbf{y} = \mathbf{x}_{\text{new}}^T\mathbf{W}\left(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{y} = \mathbf{x}_{\text{new}}^T\hat{\mathbf{b}}$.

Property 6 Using a predetermined loading weights matrix $\mathbf{W}$, the deflation order in the algorithm resulting in the orthogonlized factorizations (12) and (13) is of no importance for the residuals and predictions. A loading weights matrix $\tilde{\mathbf{W}}$ with permuted column vectors will thus give $\mathbf{X} = \tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T + \mathbf{E}_{\text{W}}$ and $\mathbf{X} = \tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T + \mathbf{E}$ respectively, with $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T = \mathbf{T}_{\text{W}}\mathbf{P}^T$ and $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T = \mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^{\mathbf{T}}$.

Proof: According to Property 3 the relation between the orthogonalized and non-orthogonalized PLS algorithms is $\mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^{\mathbf{T}} = \mathbf{T}\mathbf{W}^T$. Use of the same algorithms with the predetermined and permuted matrix $\tilde{\mathbf{W}}$ must then necessarily result in $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T = \tilde{\mathbf{T}}\tilde{\mathbf{W}}^T$. Since $\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T = \mathbf{W}\mathbf{W}^{\mathbf{T}}$ and $\tilde{\mathbf{T}}\tilde{\mathbf{W}}^T = \mathbf{T}\mathbf{W}^T$ (Property 5) it also follows that $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T\mathbf{W}\mathbf{W}^{\mathbf{T}} = \mathbf{T}\mathbf{W}^T = \mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^{\mathbf{T}}$ and thus $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T = \mathbf{T}_{\text{W}}\mathbf{P}^T$. From this follow unaltered residuals and predictions.

The OPLS algorithm Following [2], the OPLS algorithm is as follows:

1. Set $i = 1$, $\mathbf{E}_{i-1} = \mathbf{E}_0 = \mathbf{X}$, and $\mathbf{W}_{\text{ortho}}$, $\mathbf{T}_{\text{ortho}}$ and $\mathbf{P}_{\text{ortho}}$ to empty matrices
2. $\mathbf{w}_i^{\text{OPLS}} = \frac{(\mathbf{E}_{i-1})^T\mathbf{y}}{\left\|(\mathbf{E}_{i-1})^T\mathbf{y}\right\|} = \mathbf{w}_1$
3. $\mathbf{t}_i^{\text{OPLS}} = \mathbf{E}_{i-1}\mathbf{w}_i$
4. $\mathbf{p}_i^{\text{OPLS}} = \frac{(\mathbf{E}_{i-1})^T\mathbf{t}_i^{\text{OPLS}}}{\left(\mathbf{t}_i^{\text{OPLS}}\right)^T\mathbf{t}_i^{\text{OPLS}}}$
5. $\mathbf{w}_i^{\text{ortho}} = \frac{\mathbf{p}_i^{\text{OPLS}} - \mathbf{w}_i}{\left\|\mathbf{p}_i^{\text{OPLS}} - \mathbf{w}_i\right\|}$ and $\mathbf{W}_{\text{ortho}} = \left[\begin{array}{cc} \mathbf{W}_{\text{ortho}} & \mathbf{w}_i^{\text{ortho}} \end{array}\right]$
6. $\mathbf{t}_i^{\text{ortho}} = \mathbf{E}_{i-1}\mathbf{w}_i^{\text{ortho}}$ and $\mathbf{T}_{\text{ortho}} = \left[\begin{array}{cc} \mathbf{T}_{\text{ortho}} & \mathbf{t}_i^{\text{ortho}} \end{array}\right]$
7. $\mathbf{p}_i^{\text{ortho}} = \frac{\left(\mathbf{E}_{i-1}^{\text{OPLS}}\right)^T\mathbf{t}_{i+1}^{\text{ortho}}}{\left(\mathbf{t}_{i+1}^{\text{ortho}}\right)^T\mathbf{t}_{i+1}^{\text{ortho}}}$ and $\mathbf{P}_{\text{ortho}} = \left[\begin{array}{cc} \mathbf{P}_{\text{ortho}} & \mathbf{p}_i^{\text{ortho}} \end{array}\right]$
8. $\mathbf{E}_i = \mathbf{X} - \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T$
9. Let $i = i + 1$ and return to step 2 for additional orthogonal components, otherwise go to step 10
10. End.

The resulting $\mathbf{E}_i$ are the filtered $\mathbf{X}$ data, and a one component PLS factorization after removal of $i = A - 1$ components further gives

$$\mathbf{E}_{A-1} = \mathbf{t}_A^{\text{OPLS}}\left(\mathbf{p}_A^{\text{OPLS}}\right)^T + \mathbf{E}_{\text{OPLS}}. \tag{21}$$

Note that all steps give $\mathbf{w}_i^{\text{OPLS}} = \mathbf{w}_1$.

Property 7 The OPLS loading weights matrix may be found from the ordinary PLS loading weights matrix as $\mathbf{W}_{\text{ortho}} = -\mathbf{W}_{2:A}$.

Proof: From Property 6 follows that orthogonalized PLS regression with the permuted loading weights matrix $\tilde{\mathbf{W}} = \left[\begin{array}{cc} \mathbf{W}_{2:A} & \mathbf{w}_1 \end{array}\right]$ gives the same fitted response vector $\hat{\mathbf{y}}$ as with use of $\mathbf{W}$. Since the sign of a $\mathbf{w}_i$ vector has nothing to say for the products $\mathbf{t}_i\mathbf{p}_i^T$ and $\mathbf{t}_i^{\text{ortho}}\left(\mathbf{p}_i^{\text{ortho}}\right)^T$, this is true also for $\tilde{\mathbf{W}} = \left[\begin{array}{cc} -\mathbf{W}_{2:A} & \mathbf{w}_1 \end{array}\right]$. We use induction in the parameter i related to $\mathbf{W}_{\text{ortho}}$ to show that the OPLS algorithm uses $\mathbf{W}_{\text{ortho}} = -\mathbf{W}_{2:A}$.

For $i = 1$, i.e. one $\mathbf{y}$-orthogonal component, the OPLS algorithm gives

$$\mathbf{w}_1^{\text{ortho}} = \frac{\mathbf{p}_1^{\text{OPLS}} - \mathbf{w}_1}{\left\|\mathbf{p}_1^{\text{OPLS}} - \mathbf{w}_1\right\|} = \frac{\mathbf{X}^T\mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1} - \mathbf{w}_1}{\left\|\mathbf{X}^T\mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1} - \mathbf{w}_1\right\|}, \tag{22}$$

while the recursive formula for the loading weights vectors developed by Helland [7] and the prediction formula (3) give (where $\hat{\mathbf{y}}_1$ is the fitted response vector using one PLS component)

$$\begin{aligned} \mathbf{w}_2 &= \frac{\mathbf{X}^T(\mathbf{y} - \hat{\mathbf{y}}_1)}{\|\mathbf{X}^T(\mathbf{y} - \hat{\mathbf{y}}_1)\|} = \frac{\mathbf{X}^T\left(\mathbf{y} - \mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1}\mathbf{w}_1^T\mathbf{X}^T\mathbf{y}\right)}{\left\|\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1}\mathbf{w}_1^T\mathbf{X}^T\mathbf{y})\right\|} \\ &= \frac{\mathbf{w}_1 - \mathbf{X}^T\mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1}}{\left\|\mathbf{w}_1 - \mathbf{X}^T\mathbf{X}\mathbf{w}_1(\mathbf{w}_1^T\mathbf{X}^T\mathbf{X}\mathbf{w}_1)^{-1}\right\|} = -\mathbf{w}_1^{\text{ortho}}. \end{aligned} \tag{23}$$

Assuming the property to be true up to $\mathbf{w}_{i-1}^{\text{ortho}}$ we find according to the OPLS algorithm

$$\mathbf{w}_i^{\text{ortho}} = \frac{\mathbf{p}_i^{\text{OPLS}} - \mathbf{w}_1}{\left\|\mathbf{p}_i^{\text{OPLS}} - \mathbf{w}_1\right\|}, \tag{24}$$

with

$$\mathbf{E}_{i-1} = \mathbf{X} - \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T, \tag{25}$$

where $\mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T$ is the factorization of the $i-1$ removed $\mathbf{y}$-orthogonal components. From the recursive loading weights formula [7] we also find

$$\mathbf{w}_{i+1} = \frac{\mathbf{X}^T(\mathbf{y}-\hat{\mathbf{y}}_i)}{\|\mathbf{X}^T(\mathbf{y}-\hat{\mathbf{y}}_i)\|} = \frac{\mathbf{w}_1 - \frac{\mathbf{X}^T\hat{\mathbf{y}}_i}{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}}}{\left\|\mathbf{w}_1 - \frac{\mathbf{X}^T\hat{\mathbf{y}}_i}{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}}\right\|}, \tag{26}$$

where $\hat{\mathbf{y}}_i$ is the fitted response vector using a total of i components.

In order to show that $\mathbf{w}_i^{\text{ortho}} = -\mathbf{w}_{i+1}$ we finally make use of the OPLS facts that $\mathbf{T}_{\text{ortho}}^T\mathbf{y} = \mathbf{0}$ and $\mathbf{T}_{\text{ortho}}^T\hat{\mathbf{y}} = \mathbf{0}$ (see [2] for proofs), i.e. $\mathbf{E}_{i-1}^T\mathbf{y} = \left(\mathbf{X} - \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\right)^T\mathbf{y} = \mathbf{X}^T\mathbf{y}$ and $\mathbf{E}_{i-1}^T\hat{\mathbf{y}}_i = \left(\mathbf{X} - \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\right)^T\hat{\mathbf{y}} = \mathbf{X}^T\hat{\mathbf{y}}$. We then use the prediction formula (3) and the fact that OPLS gives the same predictions as ordinary PLS, and develop $\mathbf{p}_i^{\text{OPLS}}$ into (also using $\mathbf{w}_1^T\mathbf{X}^T\mathbf{y} = \mathbf{w}_1^T\mathbf{w}_1\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}} = \sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}$)

$$\begin{aligned}\mathbf{p}_i^{\text{OPLS}} &= \mathbf{E}_{i-1}^T\mathbf{E}_{i-1}\mathbf{w}_1\left(\mathbf{w}_1^T\mathbf{E}_{i-1}^T\mathbf{E}_{i-1}\mathbf{w}_1\right)^{-1} = \mathbf{E}_{i-1}^T\frac{\mathbf{E}_{i-1}\mathbf{w}_1\left(\mathbf{w}_1^T\mathbf{E}_{i-1}^T\mathbf{E}_{i-1}\mathbf{w}_1\right)^{-1}\mathbf{w}_1^T\mathbf{E}_{i-1}^T\mathbf{y}}{\mathbf{w}_1^T\mathbf{E}_{i-1}^T\mathbf{y}} \\ &= \mathbf{E}_{i-1}^T\frac{\hat{\mathbf{y}}_i}{\mathbf{w}_1^T\mathbf{E}_{i-1}^T\mathbf{y}} = \frac{\mathbf{X}^T\hat{\mathbf{y}}_i}{\mathbf{w}_1^T\mathbf{X}^T\mathbf{y}} = \frac{\mathbf{X}^T\hat{\mathbf{y}}_i}{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}},\end{aligned} \tag{27}$$

and insertion into Eq. (24) and comparison with Eq. (26) finally shows that $\mathbf{w}_i^{\text{ortho}} = -\mathbf{w}_{i+1}$.

Property 8 After the removal of $A-1$ $\mathbf{y}$-orthogonal components, the OPLS factorization (14) results in the same residual $\mathbf{E}_{\text{OPLS}} = \mathbf{E}_{\text{W}}$ and the same predictions as the original orthogonalized PLS factorization (12).

Proof: Since $\mathbf{W}_{\text{ortho}} = -\mathbf{W}_{2:A}$ the OPLS factorization is equivalent with the factorization obtained by the standard PLS NIPALS algorithm with predetermined and permuted loading weights vectors in the order $\mathbf{w}_2$, $\mathbf{w}_3$, ... , $\mathbf{w}_A$ and $\mathbf{w}_1$. From Property 6 thus follows that the residuals and the predictions are the same.

Result 1 The second similarity transformation $\mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T = \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\left(\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\right)^{-1}\mathbf{P}_{\text{ortho}}^T$ results in the transformed OPLS score matrix $\mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A} = \mathbf{T}_{2:A}$.

Proof: According to Property 3 the two factorizations $\mathbf{X} = \mathbf{T}\mathbf{W}^T + \mathbf{E}$ and $\mathbf{X} = \mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^T + \mathbf{E}$ are identical, i.e. $\mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W} = \mathbf{T}$. Using a permuted loading weights matrix $\tilde{\mathbf{W}} = \left[\begin{array}{cc}\mathbf{W}_{2:A} & \mathbf{w}_1\end{array}\right]$ we correspondlingly have $\tilde{\mathbf{T}} = \left[\begin{array}{cc}\mathbf{T}_{2:A} & \mathbf{t}_1\end{array}\right] = \tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T\tilde{\mathbf{W}}$, and that is independent of the number of components used. As the OPLS algorithm gives $\mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T$ by use of $\mathbf{X}$ and $\mathbf{W}_{\text{ortho}} = -\mathbf{W}_{2:A}$ (Property 7) in exactly the same way as we find the first $A-1$ components in $\tilde{\mathbf{T}}_{\text{W}}\tilde{\mathbf{P}}^T$, this will necessarily give

$$\mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A} = \mathbf{T}_{2:A}. \tag{28}$$

Property 9 The last OPLS component $\mathbf{t}_A^{\text{OPLS}}\left(\mathbf{p}_A^{\text{OPLS}}\right)^T$ multiplied with $\mathbf{W}\mathbf{W}^T$ becomes $\mathbf{t}_A^{\text{OPLS}}\left(\mathbf{p}_A^{\text{OPLS}}\right)^T\mathbf{W}\mathbf{W}^T = \mathbf{t}_A^{\text{OPLS}}\mathbf{w}_1^T$.

Proof: We have

$$\left(\mathbf{p}_A^{\text{OPLS}}\right)^T\mathbf{W}\mathbf{W}^T = \left(\mathbf{p}_A^{\text{OPLS}}\right)^T\left(\mathbf{w}_1\mathbf{w}_1^T + \mathbf{W}_{2:A}\mathbf{W}_{2:A}^T\right), \tag{29}$$

where

$$\left(\mathbf{p}_A^{\mathrm{OPLS}}\right)^T \mathbf{w}_1 = \frac{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1}}{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1} \mathbf{w}_1} \mathbf{w}_1 = 1 \tag{30}$$

and

$$\begin{aligned} \left(\mathbf{p}_A^{\mathrm{OPLS}}\right)^T \mathbf{W}_{2:A} &= \frac{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1}}{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1} \mathbf{w}_1} \mathbf{W}_{2:A} \\ &= \frac{\mathbf{w}_1^T \mathbf{E}_{A-1}^T}{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1} \mathbf{w}_1} \left(\mathbf{X} - \mathbf{T}_{\mathrm{ortho}} \mathbf{P}_{\mathrm{ortho}}^T\right) \mathbf{W}_{2:A} \\ &= \frac{\mathbf{w}_1^T \mathbf{E}_{A-1}^T}{\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1} \mathbf{w}_1} \left(\mathbf{T}_{2:A} - \mathbf{T}_{\mathrm{ortho}} \mathbf{P}_{\mathrm{ortho}}^T \mathbf{W}_{2:A}\right) = \mathbf{0}, \end{aligned} \tag{31}$$

where we in the final equality make use of Result 1.

Property 10 After the removal of $A-1$ $\mathbf{y}$-orthogonal components, the modified OPLS factorization (15) results in the same residual $\mathbf{E}$ and the same predictions as the modified PLS factorization (13).

Proof: Since $\mathbf{W}_{\mathrm{ortho}} = -\mathbf{W}_{2:A}$ the OPLS factorization is equivalent with the factorization obtained by the standard PLS NIPALS algorithm with predetermined and permuted loading weights vectors in the order $\mathbf{w}_2$, $\mathbf{w}_3$, ... , $\mathbf{w}_A$ and $\mathbf{w}_1$. From Property 6 and Eqs. (13) and (14) thus follows

$$\begin{aligned} \mathbf{X} &= \mathbf{T}_{\mathrm{W}} \mathbf{P}^T \mathbf{W} \mathbf{W}^T + \mathbf{E} = \left(\mathbf{T}_{\mathrm{ortho}} \mathbf{P}_{\mathrm{ortho}}^T + \mathbf{t}_A^{\mathrm{OPLS}} \left(\mathbf{p}_A^{\mathrm{OPLS}}\right)^T\right) \mathbf{W} \mathbf{W}^T + \mathbf{E} \\ &= \mathbf{T}_{\mathrm{ortho}} \mathbf{P}_{\mathrm{ortho}}^T \mathbf{W} \mathbf{W}^T + \mathbf{t}_A^{\mathrm{OPLS}} \mathbf{w}_1^T + \mathbf{E}, \end{aligned} \tag{32}$$

where the final equality making use of Property 9 results in equality with Eq. (15).

Result 2 The final modified OPLS component is identical with the first PLS+ST component, i.e. $\mathbf{t}_A^{\mathrm{OPLS}} \mathbf{w}_1^T = \mathbf{t}_1^{\mathrm{PLS+ST}} \mathbf{w}_1^T$.

Proof: When $A-1$ $\mathbf{y}$-orthogonal components are subtracted from $\mathbf{X}$, it follows from the OPLS algorithm that the remaining score vector is

$$\mathbf{t}_A^{\mathrm{OPLS}} = \left(\mathbf{X} - \mathbf{T}_{\mathrm{ortho}} \mathbf{P}_{\mathrm{ortho}}^T\right) \mathbf{w}_1 = \mathbf{E}_{A-1} \mathbf{w}_1. \tag{33}$$

Using the standard prediction formula (3) we further find

$$\hat{\mathbf{y}} = \mathbf{E}_{A-1} \mathbf{w}_1 \left(\mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{E}_{A-1} \mathbf{w}_1\right)^{-1} \mathbf{w}_1^T \mathbf{E}_{A-1}^T \mathbf{y} = \mathbf{E}_{A-1} \mathbf{w}_1 d = \mathbf{t}_A^{\mathrm{OPLS}} d, \tag{34}$$

where d is a scalar. This confirms that $\mathbf{t}_A^{\mathrm{OPLS}}$ is in the direction of $\hat{\mathbf{y}}$, which according to Property 6 is also identical with the fitted response vector using ordinary PLS regression.

From the PLS+ST factorization (5) follows $\mathbf{t}_1^{\mathrm{PLS+ST}} = q_1^{-1} \hat{\mathbf{y}}$, where q_1 is found as the first component in $\mathbf{q} = \left(\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W}\right)^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{y}$. Since $\mathbf{y}$ is orthogonal to both $\mathbf{T}_{2:A}$ (Property 1) and $\mathbf{E}$ (Property 2) we may also find q_1 by use of the PLS+ST factorization (6) and

$$\begin{aligned} \mathbf{y}^T \mathbf{X} &= \mathbf{y}^T \left(\mathbf{t}_1^{\mathrm{PLS+ST}} \mathbf{w}_1^T + \mathbf{T}_{2:A} \left(\mathbf{P}_{2:A}^{\mathrm{PLS+ST}}\right)^T + \mathbf{E}\right) = \mathbf{y}^T \mathbf{t}_1^{\mathrm{PLS+ST}} \mathbf{w}_1^T \\ &= \mathbf{y}^T q_1^{-1} \hat{\mathbf{y}} \mathbf{w}_1^T = \frac{q_1^{-1} \mathbf{y}^T \hat{\mathbf{y}}}{\sqrt{\mathbf{y}^T \mathbf{X} \mathbf{X}^T \mathbf{y}}} \mathbf{y}^T \mathbf{X}, \end{aligned} \tag{35}$$

i.e. $\frac{q_1^{-1} \mathbf{y}^T \hat{\mathbf{y}}}{\sqrt{\mathbf{y}^T \mathbf{X} \mathbf{X}^T \mathbf{y}}} = 1$ and

$$\mathbf{t}_1^{\mathrm{PLS+ST}} = q_1^{-1} \hat{\mathbf{y}} = \frac{\sqrt{\mathbf{y}^T \mathbf{X} \mathbf{X}^T \mathbf{y}}}{\mathbf{y}^T \hat{\mathbf{y}}} \hat{\mathbf{y}}. \tag{36}$$

Since $\mathbf{y}^T\mathbf{T}_{\text{ortho}} = \mathbf{0}$ we find $\mathbf{y}^T\mathbf{E}_{A-1} = \mathbf{y}^T\left(\mathbf{X} - \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\right) = \mathbf{y}^T\mathbf{X} = \sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}\mathbf{w}_1^T$, and from the $\mathbf{t}_1^{\text{PLS+ST}}$ expression (36) using $\mathbf{y}^T\mathbf{X}\mathbf{w}_1 = \sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}\mathbf{w}_1^T\mathbf{w}_1 = \sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}$ and $\hat{\mathbf{y}}$ according to Eq. (34) thus follows

$$\begin{aligned}\mathbf{t}_1^{\text{PLS+ST}} &= \frac{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}}{\mathbf{y}^T\hat{\mathbf{y}}}\hat{\mathbf{y}} = \frac{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}\mathbf{E}_{A-1}\mathbf{w}_1 d}{\mathbf{y}^T\mathbf{E}_{A-1}\mathbf{w}_1 d} \\ &= \frac{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}\mathbf{t}_A^{\text{OPLS}}}{\mathbf{y}^T\mathbf{X}\mathbf{w}_1} = \mathbf{t}_A^{\text{OPLS}}. \end{aligned} \quad (37)$$

Result 3 The first modified and then transformed OPLS loading matrix is identical with the PLS+ST loading matrix, i.e. $\mathbf{W}\mathbf{W}^T\mathbf{P}_{\text{ortho}}\left(\mathbf{W}_{2:A}^T\mathbf{P}_{\text{ortho}}\right)^{-1} = \mathbf{P}_{2:A}^{\text{PLS+ST}}$.

Proof: According to Property 3 the two factorizations $\mathbf{X} = \mathbf{T}\mathbf{W}^T + \mathbf{E}$ and $\mathbf{X} = \mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W}\mathbf{W}^T + \mathbf{E}$ are identical with $\mathbf{T}_{\text{W}}\mathbf{P}^T\mathbf{W} = \mathbf{T}$. According to Property 10 these factorizations are also identical with the modified OPLS factorization $\mathbf{X} = \mathbf{T}_{\text{ortho}}\mathbf{P}_{\text{ortho}}^T\mathbf{W}\mathbf{W}^T + \mathbf{t}_A^{\text{OPLS}}\mathbf{w}_1^T + \mathbf{E}$ and thus the transformed factorization $\mathbf{X} = \mathbf{T}_{\text{ortho}}\left(\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\right)\left(\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\right)^{-1}\mathbf{P}_{\text{ortho}}^T\mathbf{W}\mathbf{W}^T + \mathbf{t}_A^{\text{OPLS}}\mathbf{w}_1^T + \mathbf{E}$, while the PLS+ST method gives $\mathbf{X} = \mathbf{T}_{2:A}\left(\mathbf{P}_{2:A}^{\text{PLS+ST}}\right)^T + \mathbf{t}_A^{\text{PLS+ST}}\mathbf{w}_1^T + \mathbf{E}$. Since Result 1 shows that $\mathbf{T}_{\text{ortho}}\left(\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\right) = \mathbf{T}_{2:A}$, while Result 2 shows that $\mathbf{t}_A^{\text{OPLS}}\mathbf{w}_1^T = \mathbf{t}_A^{\text{PLS+ST}}\mathbf{w}_1^T$, it follows that $\left(\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}\right)^{-1}\mathbf{P}_{\text{ortho}}^T\mathbf{W}\mathbf{W}^T = \left(\mathbf{P}_{2:A}^{\text{PLS+ST}}\right)^T$.

Property 11 The modified loading matrix $\mathbf{W}\mathbf{W}^T\mathbf{P}_{\text{ortho}}$ is different from $\mathbf{P}_{\text{ortho}}$ in the last column vector only, with $\mathbf{p}_{A-1}^{\text{ortho}}$ replaced by $\left(\mathbf{p}_{A-1}^{\text{ortho}}\right)^T\mathbf{w}_1\mathbf{w}_1 - \mathbf{w}_A$.

Proof: The ordinary PLS algorithm results in an upper triangular and bi-diagonal matrix $\mathbf{P}^T\mathbf{W}$, with 1 along the main diagonal [8]. Since $\mathbf{P}_{\text{ortho}}$ in the OPLS algorithm according to Property 7 is found from $\mathbf{W}_{\text{ortho}} = -\mathbf{W}_{2:A}$ in the same way as $\mathbf{P}$ is found from $\mathbf{W}$, the matrix $\mathbf{P}_{\text{ortho}}^T\mathbf{W}_{2:A}$ must also be bi-diagonal with -1 along the main diagonal. We thus have (with $\tilde{\mathbf{p}}_i = \mathbf{p}_i^{\text{ortho}}$)

$$\begin{aligned}\mathbf{P}_{\text{ortho}}^T\begin{bmatrix}\mathbf{w}_1 & \mathbf{W}_{2:A}\end{bmatrix}\begin{bmatrix}\mathbf{w}_1^T \\ \mathbf{W}_{2:A}^T\end{bmatrix} &= \begin{bmatrix}\tilde{\mathbf{p}}_1^T\mathbf{w}_1 & -1 & \tilde{\mathbf{p}}_1^T\mathbf{w}_3 & 0 & \cdots & 0 \\ \tilde{\mathbf{p}}_2^T\mathbf{w}_1 & 0 & -1 & \tilde{\mathbf{p}}_2^T\mathbf{w}_4 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \tilde{\mathbf{p}}_{A-2}^T\mathbf{w}_1 & \vdots & & \ddots & -1 & \tilde{\mathbf{p}}_{A-2}^T\mathbf{w}_A \\ \tilde{\mathbf{p}}_{A-1}^T\mathbf{w}_1 & 0 & \cdots & \cdots & 0 & -1\end{bmatrix}\begin{bmatrix}\mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \vdots \\ \mathbf{w}_A^T\end{bmatrix} \\ &= \begin{bmatrix}\tilde{\mathbf{p}}_1^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_2^T + \tilde{\mathbf{p}}_1^T\mathbf{w}_3\mathbf{w}_3^T \\ \tilde{\mathbf{p}}_2^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_3^T + \tilde{\mathbf{p}}_2^T\mathbf{w}_4\mathbf{w}_4^T \\ \vdots \\ \tilde{\mathbf{p}}_{A-2}^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_{A-1}^T + \tilde{\mathbf{p}}_{A-2}^T\mathbf{w}_A\mathbf{w}_A^T \\ \tilde{\mathbf{p}}_{A-1}^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_A^T\end{bmatrix}. \end{aligned} \quad (38)$$

From this follows that $\tilde{\mathbf{p}}_{A-1}^T$ in $\mathbf{P}_{\text{ortho}}^T$ is replaced by $\tilde{\mathbf{p}}_{A-1}^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_A^T$, as stated. For a complete proof we must also show that for $3 \leq i \leq A$ we have $\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T - \mathbf{w}_{i-1}^T + \tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\mathbf{w}_i^T = \tilde{\mathbf{p}}_{i-2}^T$, or equivalently that $\mathbf{w}_i^T = \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T + \mathbf{w}_{i-1}^T + \tilde{\mathbf{p}}_{i-2}^T\right)$. Forming

$\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T+\mathbf{w}_{i-1}^T+\tilde{\mathbf{p}}_{i-2}^T\right)\mathbf{w}_j$ we find the following possibilities for $3\leq i\leq A$:

$$\begin{array}{rcl} j=1 & \Rightarrow & \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(\cdot)\,\mathbf{w}_j=\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1+0+\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\right)=0 \\ 1<j<i-1 & \Rightarrow & \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(\cdot)\,\mathbf{w}_j=\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(-0+0+0)=0 \\ j=i-1 & \Rightarrow & \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(\cdot)\,\mathbf{w}_j=\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(-0+1-1)=0 \\ j=i & \Rightarrow & \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(\cdot)\,\mathbf{w}_j=\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-0+0+\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)=1 \\ j>i & \Rightarrow & \left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(\cdot)\,\mathbf{w}_j=\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}(-0+0+0)=0 \end{array} \tag{39}$$

For ordinary PLS we know that $\mathbf{p}_i$ belongs to the span of $\mathbf{w}_1$, $\mathbf{w}_2$, ... $\mathbf{w}_A$, and from Property 7 and the OPLS algorithm then follows that this must be the case also for $-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1+\mathbf{w}_{i-1}+\tilde{\mathbf{p}}_{i-2}$. Since
$\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T+\mathbf{w}_{i-1}^T+\tilde{\mathbf{p}}_{i-2}^T\right)\mathbf{w}_j=1$ for $j=i$ and 0 for $j\neq i$, it finally follows that $\left(\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\right)^{-1}\left(-\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T+\mathbf{w}_{i-1}^T+\tilde{\mathbf{p}}_{i-2}^T\right)=\mathbf{w}_i^T$, and thus that $\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_1\mathbf{w}_1^T-\mathbf{w}_{i-1}^T+\tilde{\mathbf{p}}_{i-2}^T\mathbf{w}_i\mathbf{w}_i^T=\tilde{\mathbf{p}}_{i-2}^T$.

Result 4 For a single y-relevant component the relation between the post-processing PCP method [5] and PLS+ST is that $\mathbf{t}_1^{\text{PCP}}\rightarrow\mathbf{t}_1^{\text{ST}}=\mathbf{t}_A^{\text{OPLS}}$ and $\mathbf{w}_1^{\text{PCP}}\rightarrow\mathbf{w}_1$ when $\hat{\mathbf{y}}\rightarrow\mathbf{y}$, i.e. with good predictions.

Proof: PCP uses the factorization (with normalized loadings)

$$\mathbf{X}=\mathbf{t}_1^{\text{PCP}}\left(\mathbf{w}_1^{\text{PCP}}\right)^T+\mathbf{E}_{\text{PCP}}, \tag{40}$$

with $\mathbf{t}_1^{\text{PCP}}=\frac{\sqrt{\hat{\mathbf{y}}^T\mathbf{X}\mathbf{X}^T\hat{\mathbf{y}}}}{\hat{\mathbf{y}}^T\hat{\mathbf{y}}}\hat{\mathbf{y}}$ instead of $\mathbf{t}_1^{\text{PLS+ST}}=\frac{\sqrt{\mathbf{y}^T\mathbf{X}\mathbf{X}^T\mathbf{y}}}{\mathbf{y}^T\hat{\mathbf{y}}}\hat{\mathbf{y}}$ as in Eq. (36) and $\mathbf{w}_1^{\text{PCP}}=\frac{\mathbf{X}^T\hat{\mathbf{y}}}{\sqrt{\hat{\mathbf{y}}^T\mathbf{X}\mathbf{X}^T\hat{\mathbf{y}}}}$ instead of $\mathbf{w}_1=\frac{\mathbf{X}^T\mathbf{y}}{\sqrt{y^TXX^Ty}}$ as in the PLS algorithms.

References

[1] Svensson O, Kourti T, MacGregor JF. An investigation of orthogonal signal correction algorithms and there characteristics. *J. Chemometrics* 2002; **16**: 176-188.

[2] Trygg J, Wold S. Orthogonal projections to latent structures, O-PLS. *J. Chemometrics* 2002; **16**: 119-128.

[3] Martens H, Næs T. *Multivariate Calibration.* Wiley: New York, 1989.

[4] Verron T, Sabatier R, Joffre R. Some theoretical properties of the O-PLS method. *J. Chemometrics* 2004; **18**: 62-68.

[5] Langsrud Ø, Næs T. Optimised score plot by principal components of prediction. *Chemometrics Intell. Lab. Syst.* 2003; **68**: 61-74.

[6] Yu H, MacGregor JF. Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS). *Chemometrics Intell. Lab. Syst.* 2004; **73**: 199-205.

[7] Helland IS. On the structure of partial least squares regression. *Communications in statistics* 1988; **17**: 581-607.

[8] Manne R. Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics Intell. Lab. Syst* 1987; **2**: 187-97.

[9] Ergon R. PLS post processing by similarity transformation: A simple alternative to OPLS. *J. Chemometrics* 2005; **19**: 1-4